Estimating Bayesian Decision Problems with Heterogeneous Expertise*

Stephen Hansen
Universitat Pompeu Fabra†

Michael McMahon
University of Warwick‡

Sorawoot Srisuma
University of Surrey§

November 11, 2014

Abstract

We consider the recent novel two-step estimator of Iaryczower and Shum (2012), who analyze voting decisions of US Supreme Court justices. Motivated by the underlying theoretical voting model, we suggest that where the data under consideration displays variation in the common prior, estimates of the structural parameters based on their methodology should generally benefit from including interaction terms between individual and time covariates in the first stage whenever there is individual heterogeneity in expertise. We show numerically, via simulation and re-estimation of the US Supreme Court data, that the first order interaction effects that appear in the theoretical model can have an important empirical implication.

Keywords: Bayesian decision making; expertise; preferences; estimation.

JEL Codes: D72, D81, C13

*We would like to thank the editor, Badi Baltagi, and anonymous referees, as well as Greg Crawford, Carlos Velasco Rivera and Melvyn Weeks for helpful comments and discussions. Correspondence to: Sorawoot Srisuma, School of Economics, University of Surrey, Guildford, Surrey, GU2 7XH. Email: s.srisuma@surrey.ac.uk

†Other affiliations: Barcelona GSE.

‡Other affiliations: CEPR, CAGE, CEP, CfM, and CAMA.

§Corresponding author.

1 Introduction

How individuals and groups make decisions under uncertainty is important in many areas of economics and political economy, and numerous theoretical models emphasize that decision makers can differ both in terms of their knowledge of an underlying state of the world and their preferences.[1] A key challenge for taking these models to data is to estimate the decision-making parameters and understand, quantitatively, the role played by different factors in decision making. Iaryczower and Shum (2012) (hereafter IS) have proposed an empirical voting model and a novel procedure for estimating the voting behavior of US Supreme Court justices. IS consider a framework in which each justice has to vote for the Plaintiff or Defendant, based on the observed evidence and his private interpretation of the law and other specifics of the case. Specifically, each justice is allowed to differ in his ideology, or bias (π_{it}), as well as in his ability to interpret the law and the specifics of the case (θ_{it}). This decision problem is based on the theoretical voting model of Duggan and Martinelli (2001), and can be applied to other voting games (e.g. Iaryczower et al. (2013) or Hansen et al. (2014)).

IS estimate (π_{it}, θ_{it}) in two steps. In each period, a binary, unobserved state is realized; in one, the law favors the Plaintiff and in another it favors the Defendant. The first step is to estimate the probability that justices vote for the Plaintiff in both states, controlling for justice and case covariates. The second is to recover the parameters of interest by solving the structural equations imposed by the equilibrium condition of the voting game. This note proposes a simple way that can help improve their estimates. Whenever justices differ in their ability θ_{it} to perceive the state, which is typical of most interesting voting problems, the theoretical model predicts that justices will display heterogeneous responses across cases in terms of how much information they require to vote for the Plaintiff. To capture this behav-

[1] For example, see the literatures on various aspects of committee decision making (Gerling et al. 2005); career concerns (Sorensen and Ottaviani 2000, Prat 2005, Levy 2007); and political economy (Maskin and Tirole 2004, Besley 2006).

ior empirically, we propose including interaction terms in the first stage estimation. Monte Carlo simulation exercises illustrate that the interaction terms can play an important role empirically, and a re-estimation of the Supreme Court data supports the simulation results.

2 Estimation of Structural Model

This section presents the empirical model IS propose, and motivates why it may be empirically useful to explicitly allow justices with heterogeneous ability to react differently to changes in common prior beliefs that the decision should favor the Plaintiff. For brevity and notational simplicity we only consider the sincere voting version of the model.

2.1 Model

For each case t there is a common unobserved state $\omega_t \in \{0, 1\}$, unknown to every decision marker and the econometrician, that equals 1 if the law in case t favors the Plaintiff and 0 if it favors the Defendant. ω_t is drawn from a Bernoulli prior distribution with $\Pr[\omega_t = 1] = \rho_t$. Each justice i has to make a binary decision $v_{it} \in \{0, 1\}$—where 1 (0) is a vote for the Plaintiff (Defendant)—based on a private signal $s_{it} = \omega_t + \sigma_{it}\varepsilon_t$ with $\varepsilon_t \sim N(0, 1)$. An appropriate measure of expertise in this setting is $\theta_{it} = \sigma_{it}^{-1}$, which measures justice i's ability to infer the state. Justices' payoffs are state dependent and parametrized by $\pi_{it} \in (0, 1)$. All justices get a payoff of 0 if their vote matches the state. Justice i gets payoff $-\pi_{it}$ when $v_{it} = 1$ and $\omega_t = 0$, and $-(1 - \pi_{it})$ when $v_{it} = 0$ and $\omega_t = 1$. π_{it} is essentially a bias parameter that captures a justice's inclination to favor the Plaintiff: when it is close to 0 (1), the justice has a strong leaning to the Plaintiff (Defendant), while an unbiased justice has $\pi_{it} = 0.5$.

Given this setup, it can be shown that justice i chooses $v_{it} = 1$ if and only if

$$\frac{\Pr\left[\,\omega_t = 1 \mid s_{it}\,\right]}{\Pr\left[\,\omega_t = 0 \mid s_{it}\,\right]} \geq \frac{1 - \pi_{it}}{\pi_{it}}. \tag{1}$$

Bayes' Rule allows one to express

$$\ln\left(\frac{\Pr\left[\,\omega_t = 1 \mid s_{it}\,\right]}{\Pr\left[\,\omega_t = 0 \mid s_{it}\,\right]}\right) = \ln\left(\frac{\rho_t}{1 - \rho_t}\right) + \frac{2s_{it} - 1}{2\sigma_{it}^2}. \tag{2}$$

The normal distribution satisfies the Monotone Likelihood Ratio Property, which Duggan and Martinelli (2001) show implies the optimal voting rule is characterized by a threshold crossing condition. Specifically, by combining (1) and (2), it follows that $v_{it} = 1$ if and only if

$$s_{it} \geq \frac{1}{2} - \theta_{it}^{-2}\left[\ln\left(\frac{\pi_{it}}{1 - \pi_{it}}\right) + \ln\left(\frac{\rho_t}{1 - \rho_t}\right)\right] \equiv s^*\left(\theta_{it}, \pi_{it}, \rho_t\right). \tag{3}$$

Letting s_{it}^* denote $s^*\left(\theta_{it}, \sigma_{it}, \rho_t\right)$, the equilibrium probability of voting high in state ω_t is $\gamma_{it,\omega_t} \equiv 1 - \Phi\left[\theta_{it}\left(s_{it}^* - \omega_t\right)\right]$, where Φ is the normal cdf.

Expressed in this way, the voting rule (3) makes clear that justices with different expertise have heterogenous responses to changes in ρ_t. The voting rule of a justice with very high expertise will be nearly unaffected by a change in ρ_t. Since the signal is very accurate, he disregards the prior whatever its value in deciding the vote. On the other hand, the voting behavior of a justice with low expertise will be much more affected by changes in ρ_t. So, it is potentially important to allow, as a *first order effect*, for such heterogeneity in estimating voting probabilities.

The likelihood of observing the vector of votes $\mathbf{v}_t = (\mathbf{v}_{1t}, \ldots, \mathbf{v}_{nt})$ is

$$\Pr\left[\,\mathbf{v}_t\,\right] = \rho_t \prod_{i=1}^{n}\left[\gamma_{it,1}^{v_{it}}\left(1 - \gamma_{it,1}\right)^{1 - v_{it}}\right] + (1 - \rho_t)\prod_{i=1}^{n}\left[\gamma_{it,0}^{v_{it}}\left(1 - \gamma_{it,0}\right)^{1 - v_{it}}\right]. \tag{4}$$

Given $\gamma_{it,0}$ and $\gamma_{it,1}$, θ_{it} and s_{it}^* can be recovered via

$$\theta_{it} = \Phi^{-1}(1-\gamma_{it,0}) - \Phi^{-1}(1-\gamma_{it,1}) \text{ and } s_{it}^* = \frac{\Phi^{-1}(1-\gamma_{it,0})}{\Phi^{-1}(1-\gamma_{it,0}) + \Phi^{-1}(\gamma_{it,1})}. \tag{5}$$

The bias parameter π_{it} relates to all other variables in the model according to (3). Therefore one can recover (θ_{it}, π_{it}) if ρ_t, $\gamma_{it,0}$, and $\gamma_{it,1}$ are known.

2.2 Methodology

For some observable characteristics of the cases X_t and the justices Z_{it}, IS consider the following reduced form parametric terms that mimic the theoretical parameters above:

$$\rho_t(X_t;\beta) = \frac{\exp(X_t'\beta)}{1+\exp(X_t'\beta)} \tag{6}$$

$$\gamma_{it,0}(X_t, Z_{it};\zeta,\eta) = \frac{\exp(X_t'\zeta + Z_{it}'\eta)}{1+\exp(X_t'\zeta + Z_{it}'\eta)} \tag{7}$$

$$\gamma_{it,1}(X_t, Z_{it};\alpha,\delta,\zeta,\eta) = \frac{\gamma_{it,0} + \exp(X_t'\alpha + Z_{it}'\delta)}{1+\exp(X_t'\alpha + Z_{it}'\delta)}. \tag{FS:IS}$$

$\widehat{\rho}_t$, $\widehat{\gamma}_{it,0}$, and $\widehat{\gamma}_{it,1}$ can be estimated in the first stage from the maximum likelihood estimators of α, β, γ, δ, ζ, and η that maximize the natural logarithm of

$$\prod_t \left\{ \begin{array}{l} \rho_t(X_t;\beta) \prod_{i=1}^n \left[\gamma_{it,1}(X_t, Z_{it};\alpha,\delta,\zeta,\eta)^{v_{it}} (1-\gamma_{it,1}(X_t, Z_{it};\alpha,\delta,\zeta,\eta))^{1-v_{it}}\right] \\ +(1-\rho_t(X_t;\beta)) \prod_{i=1}^n \left[\gamma_{it,0}(X_t, Z_{it};\zeta,\eta)^{v_{it}} (1-\gamma_{it,0}(X_t, Z_{it};\zeta,\eta))^{1-v_{it}}\right] \end{array} \right\}. \tag{8}$$

Then in the second stage $\widehat{\theta}_{it}$ and $\widehat{\pi}_{it}$ can be obtained from solving the structural relationships in (4) and (5).

In order to allow for first order heterogenous effects for changes in s_{it}^* with respect to ρ_t, we propose an additional vector of a simple interaction terms W_{it} between elements of X_t and Z_{it} be included in the reduced form parametric terms in the first stage. More concretely, replace $\gamma_{it,0}$ and $\gamma_{it,1}$ with

$$\widetilde{\gamma}_{it,0}\left(X_t, Z_{it}, W_{it}; \zeta, \eta, \lambda\right) = \frac{\exp\left(X_t'\zeta + Z_{it}'\eta + W_{it}'\lambda\right)}{1 + \exp\left(X_t'\zeta + Z_{it}'\eta + W_{it}'\lambda\right)} \tag{9}$$

$$\widetilde{\gamma}_{it,1}\left(X_t, Z_{it}, W_{it}; \alpha, \delta, \zeta, \eta, \lambda, \xi\right) = \frac{\gamma_{it,0} + \exp\left(X_t'\alpha + Z_{it}'\delta + W_{it}'\xi\right)}{1 + \exp\left(X_t'\alpha + Z_{it}'\delta + W_{it}'\xi\right)}. \tag{FS:ALT}$$

Following the theoretical model, we expect W_{it} to play a particularly important role in empirical problems where there is a large degree of heterogeneity in justices' expertise.

3 Evaluating the importance of the interaction terms

In order to develop an intuition for how the IS methodology may generally benefit from the inclusion of interaction terms we first present some results from a small Monte Carlo study. We then replicate and re-estimate the structural parameters for the US Supreme Court voting data used in IS.

3.1 Monte Carlo

In order to test the extent to which the inclusion of interaction terms matters for the estimation of voting games, we:

1. Generate a group of 9 decision makers (the size of the Court), each making 150 independent decisions over time.

   (a) 5 members are type A with preferences π_A and expertise σ_A; 4 members are type B with preferences π_B and expertise σ_B.

   (b) We use various parameter values that are "reasonable" in the sense of being in line with estimates in IS. We examine $\pi_A = \frac{2}{3}$ and $\pi_B = \frac{1}{3}$, and

$\sigma_A = 1 - x$ and $\sigma_B = 1 + x$ for $x \in \{0, 0.05, 0.1, \ldots, 0.5\}$. So, our baseline comparisons are for eleven unique sets of parameters.[2]

2. For each unique set of π and σ values, we run 1,000 simulations. For each simulation, we generate theoretical decision data according to the following procedure:[3]

   (a) In each period t, ρ_t is drawn from $U[0.2, 0.8]$ (independent across periods).

   (b) ω_t is drawn from a Bernoulli distribution with $\Pr[\omega_t = 1] = \rho_t$.

   (c) v_{it} is drawn from a Bernoulli distribution with $\Pr[v_{it} = 1 \mid \omega_t] = \gamma_{it,\omega_t}$, as defined in section 2.

3. Given these data, we construct $X_t = (\mathbf{1}, \rho_t)$ and $Z_{it} = (\mathbf{1}, D_A)$, where D_A is a dummy variable that indicates membership of group A (and thus not actually time-varying). We use these data to estimate two separate specifications of the first-stage regressions given by (FS:IS) and (FS:ALT).

4. After we obtain estimates of first-stage coefficients, we use the structural equations (3) and (5) to recover $\widehat{\pi}_{it}$ and $\widehat{\sigma}_{it}$ for $j \in \{A, B\}$ as described above. We present as time-invariant point estimates the median values of these values across all periods.

Figure 1, which shows the percentage bias for each value of the expertise difference, summarizes the main results of the simulation exercise.[4] When expertise differences are small, the results indicate that the interaction terms do not matter much; the estimates of the parameter levels and differences are estimated reasonably

[2] As a robustness exercise we also reverse the values of the bias (i.e. $\pi_A = \frac{1}{3}$ and $\pi_B = \frac{2}{3}$) as well as consider $\pi_A = \pi_B = \frac{1}{2}$. Our findings do not change much. Numerical results are available upon request. We focus on estimation of σ rather than θ since the parameterization of the normal distribution in terms of its standard deviation is more common in many settings.

[3] Maximum Likelihood Estimation is done in R with the BFGS algorithm; code is available on request.

[4] This section focuses on the key results of the simulations, full results available on request.

accurately in specifications (FS:IS) and (FS:ALT), and neither appears to outperform the other. However, as $\sigma_A - \sigma_B$ increases, specification (FS:ALT) performs much better, especially in estimating the differences between groups, while at the same time improving in accuracy. For example, when $\sigma_A - \sigma_B = 0.6$, specification (FS:ALT) estimates $\frac{1-\pi_B}{\pi_B} - \frac{1-\pi_A}{\pi_A}$ and $\sigma_A - \sigma_B$ to 3% accuracy, whereas specification (FS:IS) displays biases of 47% and 87% respectively. Here we report the results in terms of the ratio $\frac{1-\pi}{\pi}$ since it is the key quantity for determining whether a justice votes for the Plaintiff.[5]

[Figure 1 about here.]

We also plot the complete distribution of the simulation results when $\sigma_B - \sigma_A = 0$ and when $\sigma_B - \sigma_A = 0.8$ in figure 2. With no σ differences, the results from both specifications are again very similar. But even at relatively modest expertise differences, the results show that not only does the inclusion of interaction terms ensure that the results stay anchored around the true parameters, but also that the distribution around the estimates is less dispersed too.

[Figure 2 about here.]

3.2 US Supreme Court Data

We take data from IS that contains the vote of every justice (31 in total) on every case from 1953-2008. IS run separate regressions on four subsets of cases according to the issue at stake (business, basic rights, criminal, federalism). We focus on the results for economics and basic rights cases, the two subsets IS treat as their baseline cases.

[5]The representation of this quantity as $\frac{1-\pi}{\pi}$ is a very common, but ultimately arbitrary, modelling choice. One could for example model the quantity as $\frac{1-g(\pi)}{g(\pi)}$ for any positive monotonic function g, and clearly change the magnitude of the estimated π while leaving invariant the ratio.

The first specification we run is (FS:IS), taking X_t and Z_{it} to be the same sets of variables as in IS. This replicates their results.[6] The second is (FS:ALT), including in the set of interaction terms W_{it} what appears to us to be the relevant subset of individual and meeting characteristics for influencing justices' prior beliefs.[7]

[Figure 3 about here.]

Since the effect of interaction terms only matters when there is meaningful variation in the prior ρ_t, it is important to quantify its range in the data. Figure 3 plots histograms of the estimated priors from specification (FS:IS) (the results with (FS:ALT) are very similar), and shows they range from around 0.3 to around 0.9, with a fairly dispersed distribution. This variation in the prior suggests, along with heterogeneity in justices' expertise, that interaction terms may play an important role in describing voting behaviour of judges in this dataset.

Our two specifications each produce 31 estimates (corresponding to the number of justices) of $\frac{1-\pi}{\pi}$ and σ for business and rights cases. Table 1 displays a number of summary statistics related to the distributions of these estimates. The simulation exercise above shows that not explicitly controlling for heterogeneous effects that exist across judges and cases tends to inflate estimated differences between decision makers. This is consistent with our estimates using the US Supreme Court data. As the table shows, the inclusion of the interaction terms reduces justice heterogeneity both in terms of variances and ranges. For rights case this reduction is particularly notable: the variance from the specification with interaction terms is around one sixth the value of the variance without.

[Table 1 about here.]

[6] We perform this re-estimation since IS do not report the median value of the structural parameters across all values of the fitted priors.

[7] We do not interact the mean value of *other* justices' Segal-Cover ideology or quality scores—covariates within X_{it}—with any Z_t variables, nor chief justice dummies—covariates within Z_t—with any X_{it} variables. They remain included within X_{it} and Z_t, respectively.

Finally, the radar charts in figure 4 are helpful for comparing the distributions from the two specifications more directly. Justices are ordered lowest to highest moving clockwise based on the (FS:IS) estimates. Within this disc we plot both sets of estimates. The (FS:ALT) estimates, particularly for rights cases, display notably less heterogeneity.

[Figure 4 about here.]

4 Conclusion

Given the high level of interest within economics in how individuals and groups of individuals make decisions under uncertainty, the recent two-step methodology proposed by IS provides a useful way to analyze such problems empirically. They estimate a voting model of US Supreme Court justices that accounts for voters' private information (e.g. level of expertise) and their ideological differences and this methodology can also be applied in other voting contexts.

In order to capture the main theoretical property of the model that voters with heterogeneous ability react differently to changes in the common prior belief, we propose the inclusion of interaction terms between case and justice characteristics in the first stage reduced form estimation. This should help improve the estimates of the structural parameters, especially where voters differ in their expertise. We perform some Monte Carlo studies and re-estimate the US Supreme Court data used in IS to support our estimation approach.

Finally, we end with some remarks to emphasize that we are not simply advocating making the reduced-form estimation in the first stage as flexible as possible, either by artificially including more regressors (of higher order terms) or, in the extreme, taking a completely nonparametric approach. While a more flexible specification in the first stage is appealing theoretically from the point of robustness, it may lead to more biased and imprecise estimates in the second stage, especially in finite samples. In contrast, our motivation for the inclusion of interaction terms is led by an inherent implication of voting models when voters are heterogeneous. Our numerical results show that imposing such theory-driven structure can significantly improve the structural estimates. Hence a broader message is that economic theory can be used to help inform the specification of the reduced-form component of two-step estimators in structural models.

References

Besley T. 2006. *Principled Agents?: The Political Economy of Good Government*. The Lindahl Lectures. OUP Oxford. ISBN 9780199271504.

Duggan J, Martinelli C. 2001. A Bayesian model of voting in juries. *Games and Economic Behavior* **37**: 259–294.

Gerling K, Gruner HP, Kiel A, Schulte E. 2005. Information acquisition and decision making in committees: A survey. *European Journal of Political Economy* **21**: 563–597.

Hansen S, McMahon M, Velasco Rivera C. 2014. Preferences or private assessments on a monetary policy committee? *Journal of Monetary Economics* **67**: 16 – 32.

Iaryczower M, Lewis G, Shum M. 2013. To elect or to appoint? Bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics* **97**: 230–244.

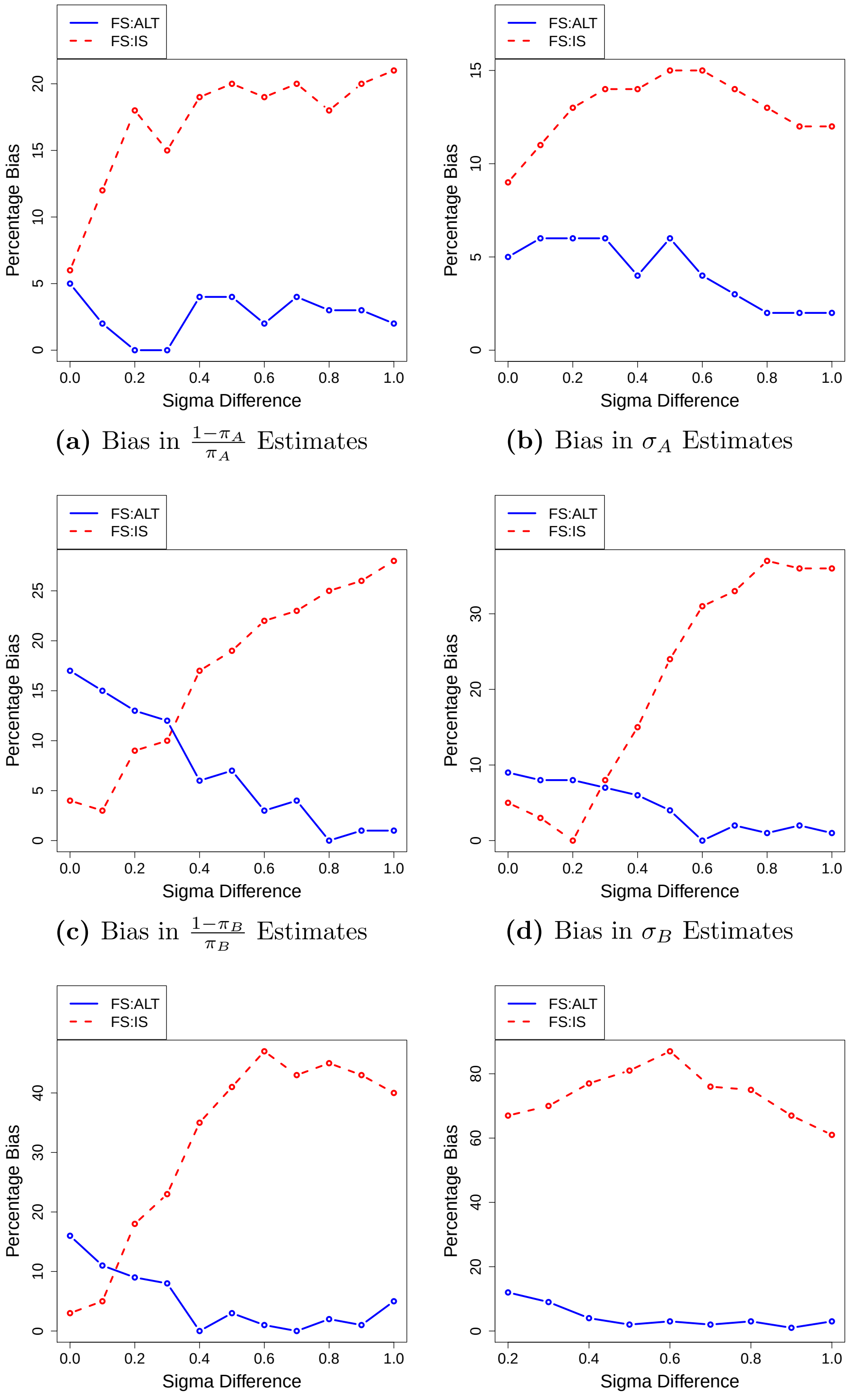
Iaryczower M, Shum M. 2012. The value of information in the court: Get it right, keep it tight. *American Economic Review* **102**: 202–37.

Levy G. 2007. Decision making in committees: Transparency, reputation, and voting rules. *American Economic Review* **97**: 150–168.

Maskin E, Tirole J. 2004. The politician and the judge: Accountability in government. *American Economic Review* **94**: 1034–1054.

Prat A. 2005. The wrong kind of transparency. *American Economic Review* **95**: 862–877.

Sorensen P, Ottaviani M. 2000. Herd behavior and investment: Comment. *American Economic Review* **90**: 695–704.



(a) Bias in $\frac{1-\pi_A}{\pi_A}$ Estimates

(b) Bias in σ_A Estimates

(c) Bias in $\frac{1-\pi_B}{\pi_B}$ Estimates

(d) Bias in σ_B Estimates

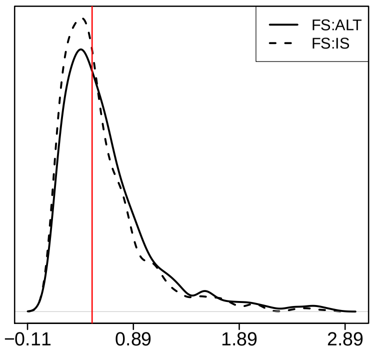
(e) Bias in $\frac{1-\pi_B}{\pi_B} - \frac{1-\pi_A}{\pi_A}$ Estimates

(f) Bias in $\sigma_B - \sigma_A$ Estimates

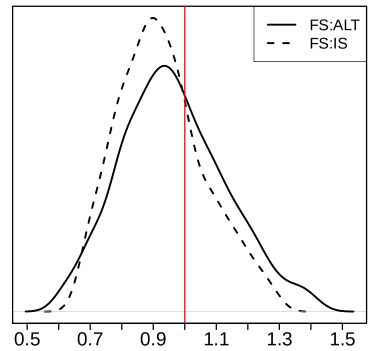
Notes: These figures plot the estimated values as a percentage of the true value (percentage bias) holding fixed $\pi_A = \frac{2}{3}$ and $\pi_B = \frac{1}{3}$ against different values of the difference $\sigma_B - \sigma_A$.

Figure 1: Percentage Biases of Estimates

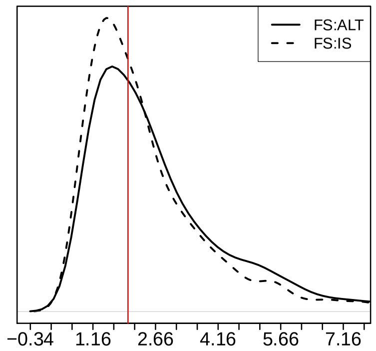
No Heterogeneity in Expertise



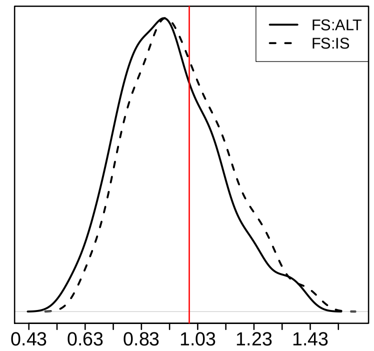
(a) Density of $\frac{1-\pi_A}{\pi_A}$ Estimates



(b) Density of σ_A Estimates

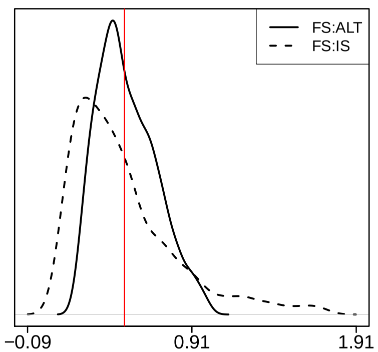


(c) Density of $\frac{1-\pi_B}{\pi_B}$ Estimates

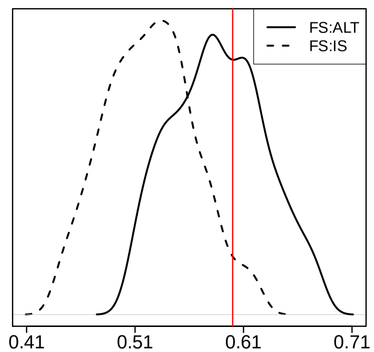


(d) Density of σ_B Estimates

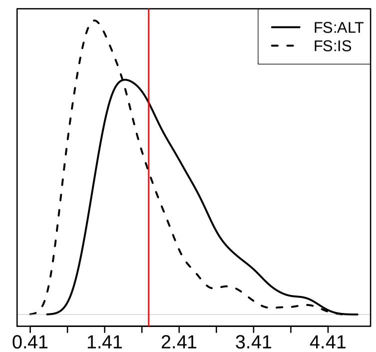
Heterogeneity in Expertise



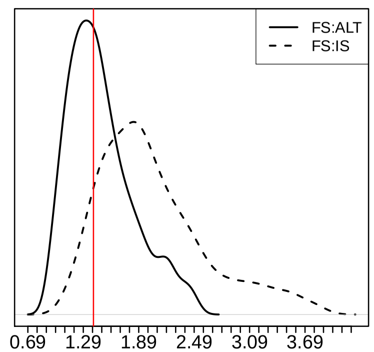
(e) Density of $\frac{1-\pi_A}{\pi_A}$ Estimates



(f) Density of σ_A Estimates



(g) Density of $\frac{1-\pi_B}{\pi_B}$ Estimates

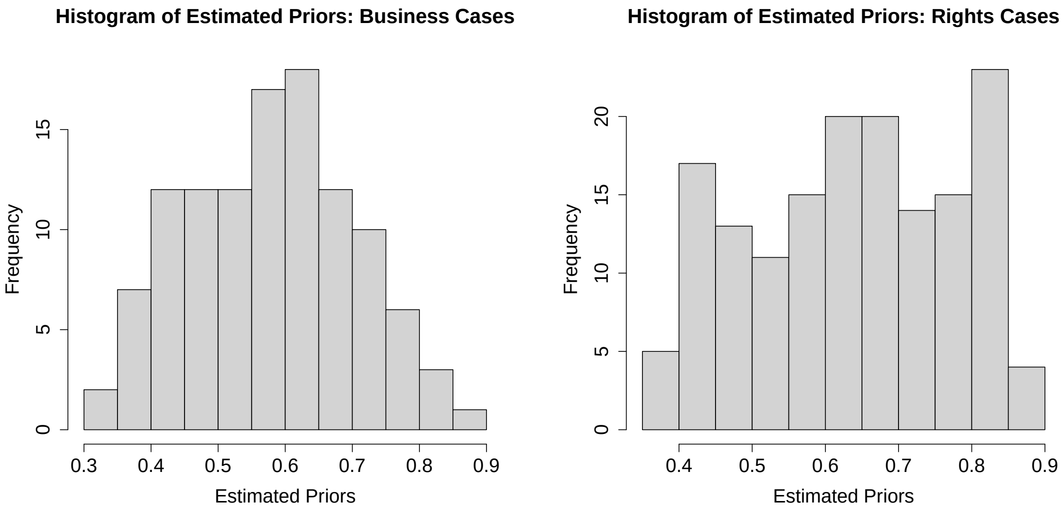


(h) Density of σ_B Estimates

Notes: These figures plot the complete distribution of the simulation results for the structural parameters of interest when $\sigma_A = 1$ and $\sigma_B = 1$ (top row) and when $\sigma_A = 0.6$ and $\sigma_B = 1.4$ (bottom row).

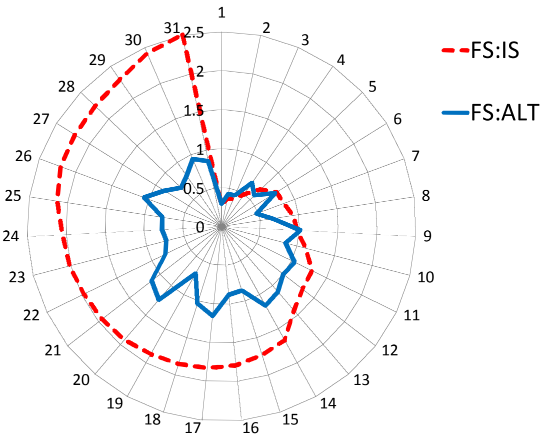
Figure 2: Densities of Estimates Without (top row) and With (bottom row) Heterogeneity in Expertise.

Figure 3: Histograms of Estimated Priors

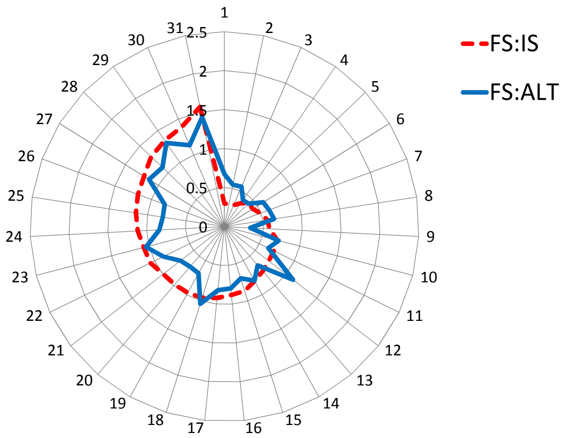


Notes: This figure plots, for business cases (left figure) and rights cases (right figure), histograms of the estimated priors ρ_t from specification (FS:IS).

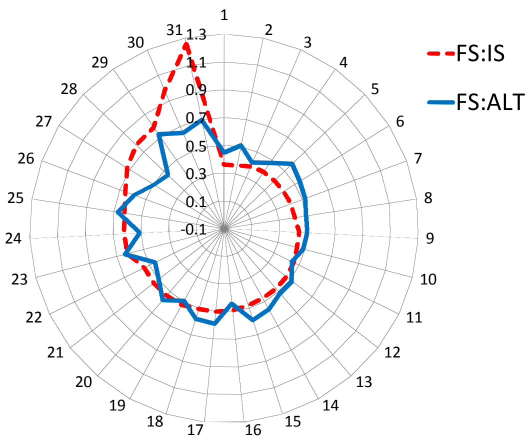
Figure 4: Radar Plots of Supreme Court Data Re-estimation Exercise



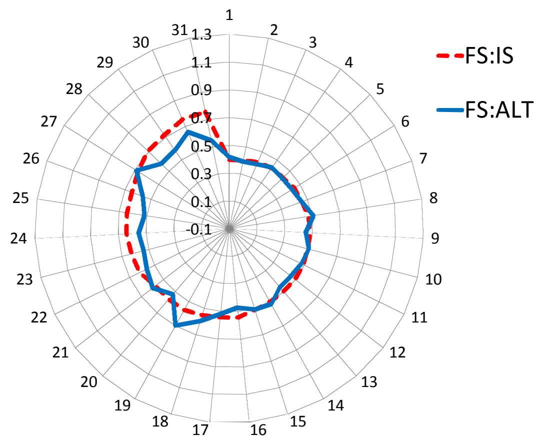
(a) $\frac{1-\pi}{\pi}$: Rights



(b) $\frac{1-\pi}{\pi}$: Business



(c) σ: Rights



(d) σ: Business

Notes: These figures show, for $\frac{1-\pi}{\pi}$ (row 1) and σ (row 2), the estimate of each Justice's parameter specification (FS:IS) along with the equivalent parameter estimated under the specification (FS:ALT). In each case, the Justices are ordered lowest to highest moving clockwise based their FS:IS estimates. Column 1 refers to Rights Cases and column 2 to Business Cases.

Table 1: Re-estimation Exercise

	Business Cases				**Rights Cases**			
	$\frac{1-\pi}{\pi}$ estimates		σ estimates		$\frac{1-\pi}{\pi}$ estimates		σ estimates	
	FS:IS	**FS:ALT**	**FS:IS**	**FS:ALT**	**FS:IS**	**FS:ALT**	**FS:IS**	**FS:ALT**
Variance	0.123	0.068	0.011	0.006	0.451	0.056	0.037	0.006
IQR	0.5249	0.3207	0.1532	0.1051	1.0802	0.3188	0.1925	0.0924
Min	0.294	0.333	0.396	0.392	0.275	0.290	0.360	0.415
Median	0.881	0.719	0.543	0.516	1.792	0.851	0.492	0.515
Max	1.567	1.431	0.759	0.699	2.509	1.240	1.255	0.726

Notes: This table shows various measures of dispersion of the distribution across judges of the estimated values of $\frac{1-\pi}{\pi}$ and σ when we specifications (FS:IS) and (FS:ALT).